Measuring Psychometric Properties of the Effective Teaching Elements for Trust School Program (TSP)

Medición de las propiedades psicométricas de los elementos de enseñanza eficaces para el programa Trust School (TSP)

Jemahliah Mohamed Salleh[1]
Universiti Kebangsaan Malaysia - MALAYSIA
jemahliah@yahoo.com
Mohd Effendi Ewan Mohd Matore[2]
Universiti Kebangsaan Malaysia – MALAYSIA
effendi@ukm.edu.my

ABSTRACT

Student-centred learning in the School Trust Program (TSP) provides a benchmark and reference centre for 21st century learning. Unfortunately, research on the psychometric properties of items for testing TSP elements is very limited. It is difficult to obtained empirical evidence for accurate, valid, and reliable measurement items. Therefore, this study aimed to test the psychometric properties of the items for TSP teaching elements through the Rasch Model. The study utilized a survey design with a fully quantitative approach. The questionnaire was adapted from the Trust School Teacher Handbook 2018, Performance Management System for Teachers.

Keywords: effective teaching instruments, Trust School Program (TSP), Rasch Model, validity, reliability.

RESUMEN

El aprendizaje centrado en el alumno en el School Trust Program (TSP) proporciona un punto de referencia y un centro de referencia para el aprendizaje del siglo XXI. Desafortunadamente, la investigación sobre las propiedades psicométricas de los artículos para probar elementos TSP es muy limitada. Es difícil obtener evidencia empírica para elementos de medición precisos, válidos y confiables. Por lo tanto, este estudio tuvo como objetivo probar las propiedades psicométricas de los elementos para elementos de enseñanza TSP a través del modelo Rasch. El estudio utilizó un diseño de encuesta con un enfoque totalmente cuantitativo. El cuestionario fue adaptado del Trust School Teacher Handbook 2018, Performance Management System for Teachers.

Palabras clave: instrumentos didácticos efectivos, Programa Trust School (TSP), Modelo Rasch, validez, confiabilidad.

1 Faculty of Education, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia.
2 Corresponding author. Center of Teaching and Learning Innovation, Faculty of Education, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

Recibido: 22/07/2019 Aceptado: 16/10/2019

1. INTRODUCTION

Teaching in the classroom should be given serious attention to ensure that the goal of effective teaching is achieved. Effective teaching can be described as a teaching that aims to achieve the intended learning outcomes or objectives or at least provide students with insight. By establishing effective teaching in the classroom, the process of transition and knowledge sharing between teachers and students can be balanced. Teachers need to understand that effective teaching can have a lasting impact on students. Asmawati et al. (2014) says that effective teaching not only gives students a real understanding of the lesson, but also develops the attitude and personality of the students. It is argued by Sakarneh (2015) that the quality of teaching is a process that can built the basis of knowledge and personal transformation of students. Teaching is an activity that requires the mastering of several skills to teach effectively and effectively. Therefore, teachers need to understand the skills used to convey knowledge and equip themselves with various teaching techniques to make teaching fun and leave a deep impression on students. Establishing the Trust School Program in Malaysia by applying teaching methods and techniques to effective teaching is expected to enhance the effectiveness of teaching in schools

The Rasch model was selected to measure the psychometric properties of the effective teaching element in the Trust School Program. The Rasch model introduced by Georg Rasch in the 1960s was a popular and growing IRT model for analysing dichotomous data into a form of scale-level data by Andrich, a partial model by Masters to facets of the model by Linacre (Sumintono & Widhiarso 2014). An individual's ability to respond to an item correctly depends on the individual's ability and difficulty. Therefore, the basis for this model is to separate the individual's ability with the quality of the instrument. This model assumes that the individual's response to an item is influenced only by the individual's ability and item difficulty (Bond & Fox 2015). The selection of the Rasch Model is because the data collected is inter-dimensional and has a small sample size. In addition, the output obtained is also easy to read, clear and easy to understand. Rasch also does not require normal data to be dispersed, so this study is well suited to using Rasch model and does not require inference analysis.

The objective of this article is to determine the validity of the items and the reliability of individual items for TSP effective teaching instrument. Item validity was studied in terms of item fit, item polarity, local independence and unidimensional analysis. Reliability is not just about studying items, it also involves individuals. In addition, the individual-separation index was also studied to identify the hierarchy of item difficulty while the individual separation index was able to distinguish between high and low performing individuals. The Rasch measurement model also analyses internal consistency values based on Cronbach's alpha values

2. PROBLEM STATEMENT

In line with the realization that effective teaching elements in TSP have not yet been covered by local scholars, studies on TSP effective teaching elements should be emphasized. Past studies have also shown a focus on effective teaching elements in the context of classroom management and teacher attitudes (Schumacher et al. 2015; Xu et al. 2015; Lans et al. 2017; Maulana et al. 2017). However, the study of TSP effective teaching elements on elements that affect the whole element of effective teaching should be emphasized. Therefore, research on these seven elements should be undertaken to obtain teaching elements that are difficult to implement by teachers. . The seven pedagogical pillars used in the TSP are teaching and delivery content, creating a positive learning environment, assessment for learning, collaborative and cooperative learning, questioning and thinking skills, differentiation and professional knowledge and reflection that need to be emphasized that the instruments used are of high quality and reliability. Several analyses were performed to assess the validity and reliability of this instrument using Rasch model approach.

3. METHODOLOGY

A quantitative survey was conducted on 203 TSP teachers at three schools in the Sate of Selangor. Questionnaires is used to obtain information from the study. The questionnaire was found to be suitable for large-sample studies and large study locations (Denscombe 2017). Sampling is based on random sampling and clustering. The sample size is derived from sample size in Rasch model as shown in Table 1.0 below. Based on the sample size table in the Rasch model, the researcher used 203 samples to obtain 99% confidence level and ± 0.5 logits.

Table 1.0Sample size in Rasch Model

Item Calibrations or person measures stable within	Confidence	Minimum sample size range	Size for most purpose
±1 logits	95%	16-36	30
±1 logits	99%	27-61	50
±0.5 logits	95%	64-144	100
±0.5 logits	99%	108-243	150

Source: *Linacre* 1994

The instruments is adapted from the Trust School Teacher Handbook 2018, Performance Management System for Teachers (LeapEd Services 2018). The instrument consists of fourteen competencies and 130 items covering seven pedagogical milestones. Table 2.0 shows the number of items by constructions and items. The instrumentation uses a five-point likert scale to get teacher feedback on effective teaching elements implemented in the classroom.

Table 2.0 Total item by construct and seven pedagogical pillars

Pillars	Pedagogical Pillars	Construct (Competency)	Total item
1	Lesson Planning, content and delivery	Plans and structures lessons effectively	16
		Uses learning objectives & success criteria	9
		Uses range of resources to support learning	6
2	Creating a positive learning environment	Establishes a physical environment	9
		Establishes an emotional environment	14
3	Assessment for learning	Uses feedback strategies	13
		Uses On-going assessment	14
		Provides opportunities for self & Peer assessment	6
4	Collaborative and cooperative learning	Uses define structures	7
		Provides opportunities for students to learn collaboratively	5
5	Questioning and thinking skills	Skilled use of questions	7
		Uses strategies to promote student thinking	5
6	Differentiation	Differentiates lessons to support individual student achievement	8
7	Professional knowledge and reflection	Applies effective learning & teaching principles	11
Total item			130

4. RESULT AND DISCUSSION

Misfit Item

The two types of compatibility statistics provided are outfit and infit mean square analysis (MNSQ) and Z scores (Z-std) that can detect whether or not an item is compatible. MNSQ's expected value is 1.00 logits. The last researcher has set the range to be followed to test the assumption that the item's MNSQ value should be within the range of 0.77 logits to 1.30 logits (Fisher, 2007), 0.5 logits to 1.5 logits, 0.6 logits to 1.4 logits (Bond & Fox, 2015). In order to test the assumption of item compatibility, researchers have chosen the MNSQ acceptance range of 0.5 to 1.5 logits. The researchers chose the MNSQ infit range of 0.5 to 1.5 logits because based on Boone et al. (2014) and Bambang & Revelation (2015) stated that items within the range are productive in measuring teachers' effective teaching level.

Z-std value indicators are used to indicate the importance of data. This is a squared mean fit statistic that estimates the theoretical mean and variance distribution. Zstd values ranging from 0.2 to 2.0 are acceptable values (Bond & Fox, 2015). In this study, the Zstd value was ignored because the MNSQ value was accepted according to Linacre (2005).

To test the assumption of item compatibility, the researcher selected the MNSQ infusion range of 0.5 to 1.5 logits (Linacre 2002; Sumintono & Widhiarso 2014; Boone et al. 2014). Findings show that the MNSQ infestation and outfit values ranged from 0.56 to 1.45. A total of 45 items were dropped from this analysis. Standard error values range from 0.18 to 0.2 for 85 items. Fisher (2007), stated an error range below 0.25 to be considered excellent. This gives the impression that the error value is small and does not interfere with the study data. In this study Zstd values were ignored based on Linacre's (2005) argument that if MNSQ is acceptable, then Zstd values can be ignored. In all, 45 items were dropped from this instrument and only 85 items matched Rasch's model.

This means that all items are productive to measure teachers' abilities and do not because the teacher to confuse the items contained in this instrument. Low and high MNSQ values can interpret the item as either overlapping the item or the item is beyond concept (Khan et al. 2013; Karim & Osman 2019). The results show that the values of 0.56 to 1.45 are within the selected range.

Unidimensionality

The testing of psychometric features in TSP effective teaching instruments emphasizes the concept of unidimensionality in the Rasch Model. In order to meet the assumption of unidimensionality as a degree of capability as emphasized by (Bejar, 1983), unidimensionality does not imply that performance on an item is due to a single process, but rather to many processes involved in answering the test item. However, as long as this process works together, the performance of the items is influenced by the same process and shape, so the concept of unidimensionality exists.

In the study, the value of raw variants explained by measurement was 45.2 percent beyond the 40% minimum assumption value in Principal Component Analysis (PCA). When the value of 45.2 percent is greater than 40 percent this value is better (Bond & Fox 2015). If it reaches 60 percent it is very good (Linacre 2005). This finding therefore concludes that the construct is good.

The study found that the item noise was 5.2 percent which is less than 10 percent and is a good value for the study. Correlation of standardized residual tests showed good results because no item exceeded the control level of 0.70. This indicates that the instrument is free from any confusion in the purpose and intent of the survey. Therefore, it can be concluded that this instrument meets the criterion of unidimensionality.

Local Independence

The range that meets local freedom requirements is a correlation value of less than 0.7 (Linacre 2012). If it is below 0.30 the value of local freedom is said to be better (Balsamo et al. 2014; Gibbons et al. 2011). In this study, the value of local freedom can provide information on whether the items are dependent on each other in the same construct or that they are not dependent on each other.

Standardized Residual Correlation test should be performed to determine if items are confusing and overlapping with each other. The analysis results show that the ten item pairs have a standardized residual correlation value of 0.62 to 0.69. This range is found to meet local conditions of freedom of correlation less than 0.70 (Linacre, 2012).

These matches have weak correlation values, that is, the items are not dependent on the items in the same construct. The findings show that the items in this instrument are not confusing and do not overlap with each other. In contrast to the study conducted by Mu'min & Hassan (2018), it shows that there are two pairs of items having local independence problems with values greater than 0.7. This means that these item pairs share several other dimensions or combine the same dimensions. If this happens, one of the items will need to be dropped.

Table 3.0 Standardized Residual Correlation

Correlation	Entry number-Competency	Entry number-competency
0.69	26 E26	27 E27
0.69	84 E84	85 E85
0.69	102 E102	103 E103
0.68	83 E83	84 E84
0.67	63 E63	64 E64
0.66	22 E22	23 E23
0.64	42 E42	43 E43
0.64	18 E18	22 E22
0.62	41 E41	42 E42
0.62	51 E51	54 E54

Reliability index and Separation Index

According to Fisher (2007), reliability values exceed 0.94 as excellent, 0.93 to 0.91 as excellent and 0.90 to 0.81 as good. An individual reliability index above 0.8 with item reliability greater than 0.9 also proved that the sample was adequate (Linacre 2012). Cronbach's alpha values were also observed with values greater than 0.80 as good.

The individual separation index indicates the number of strata identified in the sample group. The item separation index indicates the isolation of the item's difficulty level (Rahayah, 2008). The index values of the individual separation should be more than two items can be considered as good (Jones & Fox 1998).

The study found that the individual reliability index was 0.98 and the item reliability factor was 0.91. Cronbach's alpha value was 0.98. This show that these items are reliable. Individual separation index is 6.60 and item separation index is 3.10. This value can be considered as good because it exceeds two (Bond and Fox, 2015).

Meanwhile the reliability of teachers or individuals in this study is at the level of Fisher (2007). The reliability of this study instrument is that it can receive the same information or small differences even if administered repeatedly. Item and teacher reliability values closer to the value of 1.00 can give the impression that there is strong feedback from teachers as well as good instrument items (Boone et al. 2014). The value of the individual separation index in this study indicates that the instrument can distinguish high or low performance individuals. Both indexes in this study exceed the specified reference values.

Table 4.0 Summary of individual statistics

	Raw Score	Count	Measure	Model error	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	351.6	85.0	4.50	0.27	0.94	-0.3	0.93	-0.4
Standard Deviation	30.0	0.0	1.99	0.04	0.53	2.8	0.57	2.8
Max	420.0	85.0	9.45	0.47	3.30	9.3	3.35	9.6
Min	262.0	85.0	-0.51	0.19	0.05	-6.3	0.04	-6.4
Real RMSE	0.30	Adj SD	1.97	Separation	6.60	Person Reliability		0.98
Model RMSE	0.28	Adj SD	1.97	Separation	7.12	Person Reliability		0.98

Table 4.0 Summary of Item statistics

	Raw Score	Count	Measure	Model error	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	629.6	152.0	0.00	0.20	0.99	-0.1	0.93	-0.5
Standard Deviation	17.2	0.0	0.67	0.00	0.22	1.7	0.24	1.4
Max	678.0	152.0	1.79	0.20	1.45	3.2	1.50	2.6
Min	582.0	152.0	1.89	0.18	0.65	-3.0	0.56	-3.1
Real RMSE	0.21	Adj SD	0.64	Separation	3.10	Item Reliability		0.91
Model RMSE	0.20	Adj SD	0.64	Separation	3.25	Item Reliability		0.91

Item polarity

Item polarity is a basic or preliminary analysis of Rasch models. The PTMEA CORR (point measure correlation coefficient) is used to determine the polarity of an item. Polarity analysis or item alignment are indicators used to indicate that items used move in one direction as measured constructs (Siti Rahayah 2008). If the value of PTMEA Corr is high, then an item is more likely to discriminate between respondents. Negative or zero values indicate that the response to an item or respondent is inconsistent with the variables or constructs (Linacre 2005). If the PTMEA CORR value is less than 0.30 the researcher can decide whether to dismiss or improve the item. The range of polarity value of an instrument is 0.3 logits to 0.6 logits (Bond & Fox 2015). PTMEA Corr is also an early detection of construct validity.

The findings show that the Point Measure Correlation (PTMEA Corr) obtained positive values for all 85 items. Findings indicate that the polarity of items or Point Measure Correlation (PTMEA Corr) ranges from 0.39 logits to 0.79 logits. This value has a polarity value of more than 0.30 which means that all items in this instrument can distinguish or discriminate PSA teacher effective teaching items. The findings also show the positive values of PTMEA CORR for all 85 items. This shows that correlation values are good for all competencies. It can also be concluded that all items in this instrument do not conflict with the measured competencies.

The selection of the TSP's effective teaching elements is in line with the elements used by other studies, both domestically and abroad. In addition, measurement of the psychometric properties of the instrument was found to meet the assumptions of the Rasch model.

5. CONCLUSION

As a conclusion, the use of the Rasch model is to measure the psychometric properties of the items in this instrument has fulfilled all available assumptions after reference to previous Rasch model researchers and pioneers. This instrument has been found to be very effective in measuring the teaching elements of TSP. This instrument also has high reliability and reliability both in terms of item and individual.The use of the Rasch model in the measurement of psychometric properties can contribute to accurate measurement. Therefore, the Rasch model is reliable for measuring the psychometric properties of items. In this regard, the Rasch model is also a model that allows for a direct comparison between items and teacher abilities. Therefore, the Rasch model can improve existing measurements for better results.

ACKNOWLEDGMENT

This research was supported by the Universiti Kebangsaan Malaysia under grant of GGPM-2017-088 and UKM PP-FPEND-2019.

BIBLIOGRAPHIC REFERENCES

Asmawati Suhid, Mohd Rezduan Mat Jurit & Fathiyah Mohd Fakhruddin. 2014. Personaliti Pendidik Yang Ideal: Apa Kata Pelajar di Institut Pengajian Tinggi? *Journal of al-Quran and Tarbiyyah 1*(1): 1–8.

Balsamo, M., Giampaglia, G. & Saggino, A. 2014. Building a new Rasch-based self-report inventory of depression. *Neuropsychitric Disease and Treatment 10*: 153–165.

Bejar, I.I. 1983. Subject Matter Experts' Assessment of Item Statistics. *Applied Psychological Measurement 7*(3): 303–310.

Bond, T.G. & Fox, C.M. 2015a. *Applying The Rasch Model: Fundamental Measurement in the Human sciences*. Third Edition. New York: Routledge Taylor & Francis Group.

Bond, T.G. & Fox, C.M. 2015b. *Applying The Rasch model*. Third Edit. New York: Routledge New York.

Boone, W.J., Staver, J.R. & Melissa, S.Y. 2014. *Rasch Analysis in the Human Sciences*. Dordrecht: Springer.

Fisher, J.W.P. 2007. Rating scale instrument quality criteria. *Rasch Measurement Transactions 21*(1): 1095.

Jones J. A & Fox C. M. 1998. Uses of Rasch modelling in counselling psychology research. *Journal of Counselling psychology 45*(1): 30–45.

Karim, E. & Osman, K. 2019. Kesahan dan Kebolehpercayaan ujian kemahiran proses sains untuk murid sekolah rendah berdasarkan Model Pengukuran Rasch. *Jurnal Pendidikan Malaysia 1*(2018): 1–9.

Khan, A., Chien, C.-W. & Brauer, S.G. 2013. Rasch-based scoring offered more precision in differentiating patient groups in measuring upper limb function. *Journal of Clinical Epidemiology 66*(6): 681–687.

Lans, R.M. Van Der, Grift, W.J.C.M. Van De & Veen, K. Van. 2017. Developing an Instrument for Teacher Feedback : Using the Rasch Model to Explore Teachers ' Development of Effective Teaching Strategies and Behaviors. *The Journal of Experimental Education*): 1–18.

Linacre, J. 2002. Optimizing rating scale category effectiveness. *Journal of Applied Measurement 3*(1): 85–106.

Linacre, J. 2005a. *Test validity, and rasch Measurement: Construct, Content,etc.Rasch Measurement Transactions.* Chicago,Illinois: Mesa Press.

Linacre, J. 2005b. *A user's guide to WINSTEPS: Rasch Model Computer Programs*. Chicago: MESA Press.

Linacre, J. 2012. *A user's guide to WINSTEPS: Monistep Rasch Model Computer Programs*. MESA Press.

Maulana, R., Helms-lorenz M. & Grift, W. Van De. 2017. Validating a model of effective teaching behaviour of pre-service teachers. Teachers and Teaching 23(4):471-493 2017.

Martyn Denscombe. 2017. *The Good Research Guide : For Small Scale Social Reseach Projects*. 6th Editio. Leicester: Open University Press.

Mu'min & Hassan, A. 2018. Analisis validitas instrumen kompetensi dan prestasisekolah di SMP Negeri 20 Kota Batam dengan Rasch Model. *Khazanah Ilmu Berazam 1*: 1–13.

Sakarneh, M. 2015. Articulation of Quality Teaching: A Comparative Study. *Journal of Education and Training Studies 3*(1): 7–20.

Schumacher, G., Grigsby, B. & Vesey, W. 2015. Determining Effective Teaching Behaviors Through The Hiring Process. *International Journal Of Education Management 29*(1): 139–155.

Siti Rahayah Ariffin. 2008. *Inovasi Dalam pengukuran dan penilaian Pendidikan*. Bangi: Universiti Kebangsaan Malaysia.

Sumintono, B. & Widhiarso., W. 2015. *Aplikasi Pemodelan Rasch pada Assessment Pendidikan.* Cimahi: Trim Komunikata Publishing House.

Sumintono, B. & Widhiarso, W. 2014. *Aplikasi Model Rasch : Untuk Penelitian Ilmu-ilmu sosial*. Cimahi: Trim Komunikata Publishing House.

Yayasan Amir dan LeapEd Services. 2016. *Garis Panduan sekolah Amanah(GPSA)*. Edisi Pert. Kuala Lumpur: Unit Sekolah Amanah, Bahagian Pengurusan Sekolah Harian(BPSH), Kementerian Pendidikan Malaysia(KPM).